

PATENT ABSTRACTS OF JAPAN

(11)Publication number : 07-056693

(43)Date of publication of application : 03.03.1995

(51)Int.Cl.

G06F 3/06
G06F 3/06
G06F 12/08

(21)Application number : 05-201676

(71)Applicant : TOSHIBA CORP
SORD COMPUT CORP

(22)Date of filing : 13.08.1993

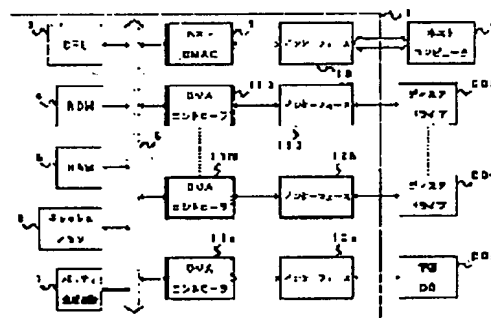
(72)Inventor : SEGAWA KIYOSHI

(54) DISK ARRAY DEVICE AND DISK WRITING CONTROLLER

(57)Abstract:

PURPOSE: To attain the efficiency and high speed of the write access of data whose size is especially small by transferring write data to a preliminary disk drive, and storing the data in it before executing a write back processing at the time of the write access.

CONSTITUTION: The write cache of storing the write data from a host computer 2 in a cache memory 6 is executed. Afterwards, the write back processing of transferring the write data from the cache memory 6 to each disk drive DD0-DDN is executed. Before the write back processing is executed, the write data are transferred to and stored in a preliminary disk driver DDS by a preliminary disk DMA controller 11S. Also, at the time of the write back processing, when the transfer of the write data from memory 6 to each disk driver DD 0-DDN is impossible, the write data stored in the preliminary disk drive DDS are read and used.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

Copyright (C); 1998,2003 Japan Patent Office

THIS PAGE BLANK (USPTO)

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号

特開平7-56693

(43) 公開日 平成7年(1995)3月3日

| (51) Int.Cl. ⁶ | 識別記号 | 庁内整理番号 | F I | 技術表示箇所 |
|---------------------------|---------|---------|-----|--------|
| G 0 6 F 3/06 | 5 4 0 | 7165-5B | | |
| | 3 0 1 J | | | |
| 12/08 | 3 2 0 | 7608-5B | | |

審査請求 未請求 請求項の数 4 O L (全 12 頁)

(21) 出願番号 特願平5-201676

(22) 出願日 平成5年(1993)8月13日

(71) 出願人 000003078

株式会社東芝

神奈川県川崎市幸区堀川町72番地

(71) 出願人 000108362

ソード株式会社

千葉県千葉市美浜区真砂5丁目20番7号

(72) 発明者 瀬川 清

千葉県千葉市真砂5丁目20番7号 ソード

株式会社内

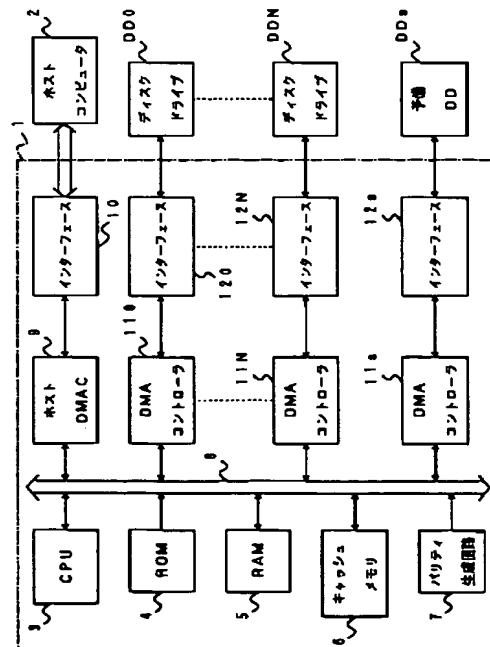
(74) 代理人 弁理士 鈴江 武彦

(54) 【発明の名称】 ディスク・アレイ装置とディスク書き込み制御装置

(57) 【要約】

【目的】 本発明の目的は、ライトキャッシュ方式を採用したディスク・アレイ装置において、コストの増大化やシステム構成の複雑化を招くことなく、特に小さいサイズのデータのライトアクセスの高速かつ効率化を図ると共に、データを確実に保護して装置の高信頼性を実現することにある。

【構成】 ホストコンピュータ2からのライトデータはキャッシュメモリ6に格納するライトキャッシュが実行される。この後に、キャッシュメモリ6から各ディスクドライブDD0～DDNにライトデータを転送するライトバック処理が実行される。このライトバック処理の実行前に、予備ディスクDMAC11sはライトデータを予備ディスクドライブDDsに転送して格納する。CPU3は、ライトバック処理が不可能な時に、キャッシュメモリ手段から各ディスクドライブへのライトデータの転送が不可能なときに、予備ディスクドライブDDsに格納したライトデータを読み出して使用するバックアップ処理を実行する。



(2)

特開平 7-56693

1

2

【特許請求の範囲】

【請求項 1】 予備ディスクドライブを含む複数のディスクドライブと、

外部から転送されたライトデータを格納するキャッシュメモリ手段と、

このキャッシュメモリ手段から前記予備ディスクドライブを除く前記各ディスクドライブに前記ライトデータを転送するライトバック処理の実行前に、前記ライトデータを前記予備ディスクドライブに転送して格納する予備ディスク制御手段とを具備したことを特徴とするディスク・アレイ装置。

【請求項 2】 予備ディスクドライブを含む複数のディスクドライブと、

外部から転送されたライトデータを格納するキャッシュメモリ手段と、

このキャッシュメモリ手段から前記予備ディスクドライブを除く前記各ディスクドライブに前記ライトデータを転送するライトバック処理の実行前に、前記ライトデータを前記予備ディスクドライブに転送して格納する予備ディスク制御手段と、

前記ライトバック処理時に、前記キャッシュメモリ手段からの前記各ディスクドライブへの前記ライトデータの転送が不可能なときに、前記予備ディスクドライブに格納した前記ライトデータを読出して使用するバックアップ制御手段とを具備したことを特徴とするディスク・アレイ装置。

【請求項 3】 予備ディスクドライブを含む複数のディスクドライブと、

外部から転送されたライトデータを格納するキャッシュメモリ手段と、

前記ライトデータの転送データ量に基づいて前記ライトデータのサイズが相対的に小さい場合に、前記予備ディスクドライブをバックアップとして使用するモードを決定するモード決定手段と、

前記モード時に、前記キャッシュメモリ手段から前記予備ディスクドライブを除く前記各ディスクドライブに前記ライトデータを転送するライトバック処理の実行前に、前記ライトデータを前記予備ディスクドライブに転送して格納する予備ディスク制御手段とを具備したことを特徴とするディスク・アレイ装置。

【請求項 4】 予備ディスクドライブを含む複数のディスクドライブを有するディスク・システムにおいて、外部から転送されたライトデータを格納するキャッシュメモリ手段と、

このキャッシュメモリ手段に格納された前記ライトデータを前記予備ディスクドライブに格納するバックアップ制御時に、前記ライトデータの識別情報、前記ライトデータを代用する無効データおよび前記ライトデータを前記予備ディスクドライブを除く前記各ディスクドライブに格納したことを指示する指示情報のそれぞれを生成す

るデータ生成手段と、

前記バックアップ制御時に、前記予備ディスクドライブの連続するアドレスに前記無効データ、前記識別情報、前記ライトデータおよび前記指示情報を逐次書込むデータ書き込み制御手段とを具備したことを特徴とするディスク書き込み制御装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】 本発明は、複数のディスクドライブを有し、ホストコンピュータからのデータを各ディスクドライブに分散して格納するディスク・アレイ装置に関する。

【0002】

【従来の技術】 従来、コンピュータシステムに使用される外部記憶装置として、RAID (Redundant Arrays of Inexpensive Disks) と呼ばれているディスク・アレイ装置が周知である。ディスク・アレイ装置は、複数のディスクドライブ (ハードディスク装置) を有し、ホストコンピュータから転送されたデータを各ディスクドライブに分散して格納する。このため、ホストコンピュータからは、あたかも 1 台のハードディスク装置をアクセスするように見える。

【0003】 ディスク・アレイ装置には、レベル 1～5 の 5 段階に分けられた異なる方式が開発されている。例えばレベル 3 の方式は、データをビット単位のインターリービング (interleaving) により各ディスクドライブに分散して格納する。また、レベル 4 の方式は、ホストコンピュータからのデータを分散だけでなく、エラー訂正用のパリティデータを格納する専用のディスクドライブを備えた方式である。

【0004】 さらに、レベル 5 の方式はレベル 4 と同様に、データをセクタ単位のインターリービングにより、各ディスクドライブに分散して格納する。レベル 5 とレベル 4 の各方式の相違は、レベル 4 がパリティデータの専用ディスクドライブを備えているに対して、レベル 5 がパリティデータを各ディスクドライブに分散して格納する点にある。

【0005】 即ち、レベル 5 の方式は、図 8 に示すように、複数のディスクドライブ DD0～DD3 にデータを論理セクタアドレス 0～11 に分散して格納すると共に、パリティデータ PD1～PD4 も分散して格納する。

【0006】 ここで、1 セクタ (論理セクタアドレス) には通常では 512 バイトのデータが格納される。ディスクドライブ DD0～DD3 では、同一物理セクタ N (N=0～3) に属する 4 つの論理アドレスセクタは 1 つのパリティグループを構成する。即ち、パリティデータ PD1～PD4 はそれぞれ、パリティグループを構成する各論理アドレスセクタの各データと同一バイト位置

(3)

特開平7-56693

3

に書込まれる。例えばパリティデータPD1は、論理アドレスセクタ0～2とパリティグループを構成する。

【0007】このようなレベル5の方式では、1台のディスクドライブ（例えばDD0）が故障しても、残りの3台のディスクドライブ（DD1～DD2）のデータとパリティデータPD1を利用することにより、故障したディスクドライブ（例えばDD0）のデータを修復することができる。したがって、単独のハードディスク装置と比較して、高い信頼性を確保することができる。

【0008】しかしながら、2台以上のディスクドライブに故障が発生すると、前記の方法ではデータの修復は不可能である。このため、図8に示すように、予備ディスクドライブDDsを用意して、1台目のディスクドライブの故障が発生したときに、前記の方法でデータを修復し、この修復データを予備ディスクドライブDDsにコピーする方式が開発されている。この方式では、当然ながら予備ディスクドライブDDsは、1台目のディスクドライブに故障が発生するまでアクセスされない。

【0009】一方、レベル5の方式はレベル3の方式と比較して、画像データ等の転送データ量（データのサイズ）が大きいデータよりも、小さいサイズのデータのアクセスに有効である。レベル3の方式は、1セクタ分のデータをアクセスするときに、複数のディスクドライブの全てをアクセスする必要がある。このため、データ転送速度は高速となるため大きいサイズのデータ転送には都合がいいが、小さい単位のデータアクセスでは効率が低下する。

【0010】しかし、レベル5においても、小さいサイズのデータをライトアクセスするときには、パリティデータの書換えを必要とするため性能が低下する。このため、特にライトアクセス時に、キャッシュメモリを使用するライトキャッシュ方式を採用した装置が開発されている。即ち、例えば1セクタ分のライトアクセスでは、書換える前のデータ（旧ユーザデータ）とそれに対応するパリティデータ（旧パリティデータ）をそれぞれリードアクセスする。これらの旧ユーザデータ、旧パリティデータおよび新たなライトデータ（新ユーザデータ）から新たなパリティデータ（新パリティデータ）を生成する。したがって、1セクタ分のライトアクセスには、2リードアクセス、新ユーザデータのライトアクセスおよび新パリティデータのライトアクセスの計4アクセスが必要となる。

【0011】このような欠点を解消するために、レベル5に前記のライトキャッシュ方式を採用した装置がある。この方式は、ホストコンピュータからライトデータ（新ユーザデータ）を受信すると、キャッシュメモリに格納してライトコマンドを終了する。この後に、ディスクドライブのアイドル時（非アクセス時）に、ライトデータと新パリティデータを指定の各セクタに書込むライトバック処理を実行する。

4

【0012】

【発明が解決しようとする課題】従来のレベル5のディスク・アレイ装置において、ライトキャッシュ方式を採用することにより、小さいサイズのデータのライトアクセスを高速かつ効率的に行なうことが可能となる。しかしながら、この方式では、キャッシュメモリからデータをディスクドライブに転送するライトバック処理が必要となる。このライトバック処理の実行終了前に、キャッシュメモリに故障が発生したり、またはキャッシュメモリの電源供給が停止して、キャッシュメモリに格納されたデータが破壊または消去するような事態が発生する可能性がある。このような事態が発生すれば、ライトバック処理は不可能となる。このため、単にライトキャッシュを採用した方式では、データ保護が不完全であり、装置の信頼性が低いという問題がある。

【0013】信頼性を高めるためには、キャッシュメモリやディスク・アレイ・コントローラの2重化、またはキャッシュメモリに不揮発性メモリや無停電電源装置を使用するなどの方法が考えられる。しかし、このような方法を採用すると、装置全体のコストが増大し、かつシステム構成の複雑化を招く欠点がある。

【0014】本発明の目的は、ライトキャッシュ方式を採用したディスク・アレイ装置において、特に小さいサイズのデータのライトアクセスの高速かつ効率化を図ると共に、コストの増大化やシステム構成の複雑化を招くことなく、データを確実に保護して装置の高信頼性を実現することにある。

【0015】

【課題を解決するための手段】本発明は、例えばレベル5のディスク・アレイ装置において、予備ディスクドライブを含む複数のディスクドライブ、ライトデータを格納するキャッシュメモリ手段およびライトアクセス時のライトバック処理の実行前に、ライトデータを予備ディスクドライブに転送して格納する予備ディスク制御手段を有する装置である。

【0016】さらに、本発明はライトバック処理時に、キャッシュメモリ手段から各ディスクドライブへのライトデータの転送が不可能なときに、予備ディスクドライブに格納したライトデータを読出して使用するバックアップ制御手段を有する装置である。

【0017】

【作用】本発明では、ホストコンピュータからのライトデータはキャッシュメモリ手段に格納するライトキャッシュが実行される。この後に、キャッシュメモリ手段から各ディスクドライブにライトデータを転送するライトバック処理が実行される。このライトバック処理の実行前に、予備ディスク制御手段はライトデータを予備ディスクドライブに転送して格納する。

【0018】さらに、本発明では、バックアップ制御手段はライトバック処理時に、キャッシュメモリ手段から

(4)

特開平7-56693

5

各ディスクドライブへのライトデータの転送が不可能なときに、予備ディスクドライブに格納したライトデータを読出して使用する。

【0019】

【実施例】以下図面を参照して本発明の実施例を説明する。図1は同実施例に係わるディスク・アレイ装置（RAID装置）の要部を示すブロック図、図2乃至図4および図7は同実施例の動作を説明するためのフローチャート、図5および図6は同実施例に係わる予備ディスクドライブの記憶内容を説明するための概念図である。

【0020】本装置は、図1に示すように、複数のディスクドライブDD0~DDNと1台の予備ディスクドライブDDs、および各ディスクドライブを制御するディスク・アレイ・コントローラ（以下単にコントローラと称する）1を有する。コントローラ1は、レベル5の方式により各ディスクドライブDD0~DDNを制御する。即ち、コントローラ1は、ホストコンピュータ2から転送されるライトデータを、セクタ単位のインターリーブングにより各ディスクドライブDD0~DDNに分散して格納する。さらに、コントローラ1は、ライトデータに対応するパリティデータを生成して各ディスクドライブDD0~DDNに分散して格納する。ここで、ホストコンピュータ2との間で転送するリードデータまたはライトデータを、パリティデータと区別する場合にはユーザデータと称する。

【0021】コントローラ1は、マイクロプロセッサ（CPU）3、読出し専用メモリ（ROM）4、リード／ライトメモリ（RAM）5、キャッシュメモリ6、パリティ生成回路7および内部バス8を有する。CPU3は、ROM4に予め格納されたプログラムにより、各ディスクドライブDD0~DDNのリード／ライト制御、本発明に係わるライトバック処理およびバックアップ処理を実行する。RAM5はCPU3のワークメモリとして使用される。キャッシュメモリ6は、リード／ライトデータを一時的に格納するための高速バッファメモリである。パリティ生成回路7は、ホストコンピュータ2からのライトデータに対応するパリティデータを生成するための回路である。

【0022】さらに、コントローラ1は、ホストDMAコントローラ（ホストDMAC）9、インターフェース10、ディスクDMAコントローラ（ディスクDMAC）110~11N、インターフェース120~12N、予備ディスクDMAコントローラ（予備DMAC）11sおよびインターフェース12sを有する。ホストDMAC9は、ホストバス13と内部バス8間のデータ転送を制御し、本発明に係るライトアクセス時にホストバス13からのライトデータをキャッシュメモリ6に転送する。インターフェース10はホストバス13に接続されて、ホストコンピュータ2とのデータ交換を実行する。

6

【0023】ディスクDMAC110~11Nはそれぞれのインターフェース120~12Nを介して、各ディスクドライブDD0~DDNと内部バス8との間のデータ転送を制御する。予備DMAC11sはインターフェース12sを介して、予備ディスクドライブDDsと内部バス8との間のデータ転送を制御する。

【0024】次に、同実施例の動作を説明する。

（基本動作）まず、図2のフローチャートを参照して、レベル5のディスク・アレイ装置の基本的動作を説明する。ホストコンピュータ（ホストCPU）2からリード／ライト（R/W）コマンドがホストバス13を介して転送されると、ホストDMAC9は内部バス8を通じてR/WコマンドをRAM5に転送する（ステップS1、S2）。CPU3は、ホストCPU2からのR/Wコマンドを解釈してディスクコマンドに変換する（ステップS3）。ディスクコマンドは、各ディスクドライブDD0~DDNのリード／ライト制御を実行するためのコマンドである。

【0025】CPU3は変換したディスクコマンドをRAM5に格納する。ディスクDMAC110~11Nは、RAM5からリード／ライト動作を指示するディスクコマンドを各ディスクドライブDD0~DDNに転送する。

（リードアクセス）ディスクコマンドがリードコマンドであれば（ステップS4のYES）、ディスクDMAC110~11Nはリードコマンドを各ディスクドライブDD0~DDNに転送する（ステップS6）。各ディスクドライブDD0~DDNは、指定された論理セクタアドレスのデータを読出すリードアクセスを実行する（ステップS7）。ディスクDMAC110~11Nは、読出したリードデータをキャッシュメモリ6に転送する（ステップS8）。

【0026】ホストDMAC9は、転送順序をチェックしながら、キャッシュメモリ6に格納されたリードデータをインターフェース10を介してホストCPU2に転送する（ステップS9）。ここで、例えばディスクドライブDD0のディスクにディフェクトが存在したり、データの破壊が発生したりして、データの読出し動作が不可能な場合には、CPU3はデータの修復処理を実行する（ステップS10のNO、S11）。即ち、CPU3は、ディスクDMAC110を介して、リードできないデータの同一パリティグループに属するユーザデータとパリティデータを読出し、パリティ計算に基づいて元のデータを修復してキャッシュメモリ6に格納する。ホストDMAC9は、キャッシュメモリ6から修復したデータをリードデータとしてホストCPU2に転送する。CPU3は、正常にリードコマンドが終了すると、正常終了ステータスをホストCPU2に送信する。

【0027】CPU3は、前記のように例えばディスクドライブDD0に障害が発生すると、ディフェクトの交

(5)

特開平7-56693

7

替処理を実行したり、または故障したディスクドライブDD0の代わりに予備ディスクドライブDDsを代替する。この代替処理では、故障したディスクドライブDD0に格納した全データを修復して、予備ディスクドライブDDsに格納する。以後、CPU3は予備ディスクドライブDDsを正規のディスクドライブDD0として使用する。

【0028】ホストCPU2からのコマンドがライトコマンドであれば（ステップS4のNO）、CPU3は各ディスクドライブDD0～DDNに対するライトアクセスを実行する（ステップS5）。同実施例では、ホストDMAC9は、ホストCPU2からのライトデータをキャッシュメモリ6に格納するライトキャッシュを実行する。

【0029】以下図3のフローチャートを参照して、同実施例に係わるライトアクセスの動作を説明する。同実施例では、CPU3はライトデータのサイズを判定し、比較的大きいサイズのライトデータあれば、通常のライトアクセス動作を実行する。また、比較的小さいサイズのライトデータあれば、本発明に係わるライトアクセス動作を実行する。

【0030】即ち、CPU3は、ホストDMAC9により転送されたライトコマンドにより、ライトデータの転送データ量を認識して、ライトデータのサイズを判定する（ステップS12）。このとき、CPU3はキャッシュメモリ6の使用可能な格納容量も検出する。

（大きいサイズのライトアクセス）ライトデータが例えば画像データのように大きいサイズのデータであれば（ステップS13のNO）、ホストDMAC9は、ホストCPU2からのライトデータをキャッシュメモリ6に転送する（ステップS14）。CPU3は、パリティ生成回路7を使用して、ライトデータに対応するパリティデータを生成してキャッシュメモリ6に格納する（ステップS15）。即ち、CPU3は、ライトデータの書換え対象である旧ユーザデータと旧パリティデータを読み出し、これらの旧データに基づいて新たなパリティデータを生成する。この場合、キャッシュメモリ6は単なるバッファメモリとして使用される。

【0031】ディスクDMAC110～11Nは、キャッシュメモリ6からライトデータと新パリティデータを、各ディスクドライブDD0～DDNに転送する。各ディスクドライブDD0～DDNは、ライトデータと新パリティデータを、ディスクコマンドにより指定された論理セクタアドレスに格納する（ステップS17）。CPU3は、各ディスクドライブDD0～DDNの全てが正常にライトコマンドを終了すると、正常終了ステータスをホストCPU2を送信する。

【0032】このようなライトアクセス時においても、ディスクドライブDD0～DDNに障害が発生すると、前記のリードアクセスの場合と同様に、CPU3はディ

8

フェクトの交替処理を実行したり、または故障したディスクドライブの代わりに予備ディスクドライブDDsを代替する。この代替処理では、故障したディスクドライブDD0に格納した全データを修復して、予備ディスクドライブDDsに格納する。以後、CPU3は予備ディスクドライブDDsを正規のディスクドライブとして使用する。

（小さいサイズのライトアクセス）ライトデータが小さいサイズのデータであれば（ステップS13のYES）、ホストDMAC9は、ホストCPU2からのライトデータをキャッシュメモリ6に転送する（ステップS18）。次に、本発明では、予備DMAC11sはインターフェース12sを介して、キャッシュメモリ6に格納されたライトデータを予備ディスクドライブDDsに転送する（ステップS19）。これにより、予備ディスクドライブDDsはライトデータを格納し、バックアップすることになる（ステップS20）。このとき、予備DMAC11sは、ホストDMAC9により転送されるライトデータをそのまま予備ディスクドライブDDsに転送してもよい。

【0033】この後、CPU3は、各ディスクドライブDD0～DDNのアイドル時（非アクセス時）に、バックグラウンドジョブにより、キャッシュメモリ6に格納されたライトデータを各ディスクドライブDD0～DDNに格納するライトバック処理を実行する（ステップS21）。

【0034】ライトバック処理では、図4に示すように、CPU3は、パリティ生成回路7を使用して、ライトデータに対応するパリティデータを生成してキャッシュメモリ6に格納する（ステップS22）。ディスクDMAC110～11Nは、キャッシュメモリ6からライトデータと新パリティデータを、各ディスクドライブDD0～DDNに転送する（ステップS23）。各ディスクドライブDD0～DDNは、ライトデータと新パリティデータを、ディスクコマンドにより指定された論理セクタアドレスに格納する（ステップS24）。

【0035】CPU3は、各ディスクドライブDD0～DDNの全てが正常にライトコマンドを終了すると、正常終了ステータスをホストCPU2を送信する（ステップS25のYES）。このようなライトバック処理が正常に終了するまで、予備ディスクドライブDDsにはライトデータがバックアップされている。

【0036】ここで、ライトバック処理が正常に終了する前に、キャッシュメモリ6に故障が発生したり、またはキャッシュメモリ6の電源供給が停止して、キャッシュメモリ6に格納されたデータが破壊または消去するような事態が発生したとする（ステップS25のNO）。予備DMAC11sは、CPU3の指示に応じて、予備ディスクドライブDDsにバックアップしてあるライトデータをリードする（ステップS26）。ディスクDM

(6)

特開平7-56693

9

10

AC110~11Nは、予備ディスクドライブDDsからリードされたライトデータを、指定された論理セクタアドレスにライトバックする（ステップS27）。このとき、CPU3は、パリティ生成回路7を使用して、ライトデータに対応するパリティデータを新たに生成する。

【0037】このようにして、小さいサイズのライトデータを書込む場合には、キャッシュメモリ6に格納すると共に、予備ディスクドライブDDsにバックアップする。このライトキャッシュとバックアップ処理により、ライトコマンドは終了となる。この後、ホストCPU2からのリード／ライトアクセスのない各ディスクドライブDD0~DDNのアイドル時に、キャッシュメモリ6から各ディスクドライブDD0~DDNに対して、ライトデータのライトバック処理が実行される。

【0038】予備ディスクドライブDDsにバックアップされたライトデータは、ライトバック処理が終了するまで保存されている。ライトバック処理が終了する前に、キャッシュメモリ6に故障または電源供給の停止の事態が発生して、ライトバック処理が不可能になったときに、予備ディスクドライブDDsにバックアップされたライトデータが利用される。したがって、結果的にライトデータは常に確実に保護されることになり、装置の信頼性を高めることが可能となる。

【0039】ところで、ディスクドライブDD0~DDNに障害が発生した場合には、前記の大きいサイズのライトデータのアクセス時と同様に、予備ディスクドライブDDsを利用してデータの修復処理を実行する。このとき、まずキャッシュメモリ6に格納されたライトデータの全てをライトバック処理する。この後に、故障したディスクドライブに格納した全データを修復し、予備ディスクドライブDDsに格納する。以後、ライトキャッシュを禁止して、キャッシュメモリ6を単なるバッファメモリとして使用し、CPU3は予備ディスクドライブDDsを正規のディスクドライブとして使用する。

（予備ディスクドライブDDsの動作）次に、本発明に係わる予備ディスクドライブDDsの動作を、図5乃至図7を参照して具体的に説明する。

【0040】まず、CPU3は予備ディスクドライブDDsをライトアクセスするときに、シーケンシャル・ライトコマンドを出力する。即ち、図5に示すように、アクセスされるセクタアドレスは初期値（アドレス0）からシーケンシャルに増加し、順次セクタ単位のプロックデータが各アドレスに格納される。同実施例では、キャッシュメモリ6に格納されたライトデータが全て各ディスクドライブDD0~DDNにライトバックされたときに、予備ディスクドライブDDsのセクタアドレスは初期値に復帰する。また、1回のライトコマンドにより、ライトできるブロック数はコマンド、ステータスのやりとりによるオーバーヘッドを十分に無視できる程度の大き

い値とする。

【0041】図7のフローチャートに示すように、CPU3は、シーケンシャル・ライトコマンドを出力して、予備ディスクドライブDDsにバックアップすべきライトデータをライトする制御を行なう（ステップS30）。ここで、前記の大きいサイズのライトデータのライトアクセスの場合には、バックアップすべきライトデータはないため、CPU3はヌルブロック（ダミーデータ）をRAM5に生成する（ステップS31のNO、S32）。予備DMAC11sは、CPU3により生成されたヌルブロックを転送し、図5に示すように、例えばセクタアドレス0にヌルブロックを格納する（ステップS33）。

【0042】一方、ホストCPU2からバックアップすべき小さいサイズのライトデータが転送されると、前記のように、ライトデータはキャッシュメモリ6に格納されて、ライトバック処理が実行される（ステップS31のYES、S34）。CPU3は、1ブロック長のヘッダブロックをRAM5に生成する（ステップS35）。予備DMAC11sは、RAM5から予備ディスクドライブDDsにヘッダブロックを転送し、例えばセクタアドレス2に格納する（ステップS36）。続いて、予備DMAC11sは、キャッシュメモリ6から予備ディスクドライブDDsにライトデータ（ユーザデータ）を転送し、例えばセクタアドレス3~6に格納する（ステップS37）。

【0043】ここで、予備DMAC11sは、ライトデータの転送時に予備ディスクドライブDDsのディスク回転待ちの状態が発生しないように、ヌルブロックとヘッダブロックを連続的に転送する。CPU3は、ライトデータの転送後にヌルブロックを転送し、ライトバック処理に移行する（ステップS38、S39のYES）。即ち、キャッシュメモリ6に格納されたライトデータは、各ディスクドライブDD0~DDNにライトバックされる（ステップS40）。ライトデータがライトバックされると、CPU3はコンプリートブロック（Cブロック）をRAM5に生成する（ステップS41）。予備DMAC11sは、RAM5から予備ディスクドライブDDsにコンプリートブロックを転送し、例えばセクタアドレス8に格納する（ステップS42）。

【0044】ここで、ヌルブロック、ヘッダブロックおよびCブロックは、図6に示すように、それぞれ512バイトのデータ長からなる。ヌルブロックは、1バイトのヌルコード（例えば00h）と1バイトのバックアップIDコード（以下単にIDと称する）を有する。残りのバイトは無効データからなる。ヌルコードは、そのブロック（セクタ）にはユーザデータが存在しないことを示すコードである。

【0045】ヘッダブロックは、1バイトのヘッダコード（例えば01h）、1バイトのID、1バイトのコマ

(7)

特開平7-56693

11

ンドタグおよびホストCPU2から転送されたライトコマンドを有する。ヘッダコードは、ユーザデータの先頭部であることを示すコードである。Cブロックは、1バイトのコンプリートコード（例えば02h）、1バイトのIDおよび1バイトのコマンドタグを有する。IDとコマンドタグはライトバック処理が不可能な場合に、ライトバックできなかったユーザデータ（ライトデータ）を検索するために使用される。IDは、予備ディスクドライブDDsにシーケンシャル・ライトが実行されて、アドレスが初期値に復帰したときに「1」だけインクリメントされる。ライトバックされなかったユーザデータを検索するときは、予備ディスクドライブDDsのセクタアドレス（初期値から）をシーケンシャルにリードして、初期値のアドレスの内容（マルチブロック又はヘッダブロック）のIDとは異なるIDが見付るまでリードする。このリードしている期間に、ヘッダブロックに対応するCブロックをチェックし、その対応するCブロックが見付らないヘッダブロックがあれば、それがライトバックされなかったライトコマンドである。そして、ヘッダブロックのつぎのブロックからはユーザデータが書き込まれているので、このユーザデータを読出して各ディスクドライブDD0~DDNに転送してデータ修復処理を行なう。

【0046】このようにして、小さいサイズのライトデータを各ディスクドライブDD0~DDNに転送するときには、予備ディスクドライブDDsをデータのバックアップ用として使用する。予備ディスクドライブDDsでは、ライトデータ（ユーザデータ）の先頭にヘッダブロックが格納されて、ライトバック処理の終了後にCブロックが格納される。ライトデータがないときでも、CPU3はライトコマンドを予備ディスクドライブDDsに出力している。この場合には、予備DMAC11sは、CPU3が生成したマルチブロックを繰り返し転送する。

【0047】ホストCPU2からバックアップすべきライトデータが転送されて、キャッシュメモリ6に格納されると、CPU3はヘッダブロックを生成して、予備DMAC11sを介して予備ディスクドライブDDsに転送する。この後に、キャッシュメモリ6に格納されたライトデータを予備ディスクドライブDDsに転送し、再度マルチブロックの転送に戻る。そして、ライトバック処理が終了すると、CPU3はCブロックを生成して、予備DMAC11sを介して予備ディスクドライブDDsに転送する。

【0048】このような予備ディスクドライブDDsを使用した方式であれば、予備ディスクドライブDDsを使用したバックアップ処理では、シーケンシャル・ライトアクセスであるため、ドライブDDsのヘッドのシーク時間およびディスクの回転待ち時間はほぼ0である。この予備ディスクドライブDDsを使用したバックアップ

12

方式とライトキャッシュ方式を併用すれば、ライトキャッシュ方式による高速アクセスとバックアップ方式によるデータの保護機能を実現することができる。したがって、結果的には二重化ライトキャッシュ付き無停電電源装置を備えたディスク・アレイ装置と同等の性能と信頼性を、二重化なしライトキャッシュ付き無停電電源装置なしの装置とほぼ同等のコストにより得ることができる。

【0049】なお、予備ディスクドライブDDsに故障が発生した場合には、他のディスクドライブDD0~DDNの故障時と同様に、コントローラ1はホストCPU2またはオペレータに故障発生のアラームを発生する。このとき、CPU3はキャッシュメモリ6の内容を全てライトバックし、予備ディスクドライブDDsの修理または交換が終了するまでキャッシュメモリ6を使用したライトキャッシュ動作を停止する。

【0050】また、ライトバック処理が終了する前に、装置の電源がオフされた場合に備えて、電源がオンされた直後にCPU3は予備ディスクドライブDDsのヘッダブロックとCブロックチェックを行なうようにしてもよい。但し、ライトバック処理が終了する前に、キャッシュメモリ6と予備ディスクドライブDDsの両方が同時に故障する確率はほとんど0と考えられる。

【0051】

【発明の効果】以上詳述したように本発明によれば、ライトキャッシュ方式を採用したディスク・アレイ装置において、予備ディスクドライブをライトデータのバックアップ用に利用することにより、キャッシュメモリやディスク・アレイ・コントローラの2重化、またはキャッシュメモリに不揮発性メモリや無停電電源装置を使用などの方法を採用することなく、小さいサイズのデータのライトアクセスの高速性と確実なデータ保護を実現することができる。したがって、コストの増大化やシステム構成の複雑化を招くことなく、特に小さいサイズのデータのライトアクセスの高速かつ効率化を図ると共に、データを確実に保護して装置の高信頼性を実現できることになる。

【図面の簡単な説明】

【図1】本発明の実施例に係わるディスク・アレイ装置の要部を示すブロック図。

【図2】同実施例の動作を説明するためのフローチャート。

【図3】同実施例の動作を説明するためのフローチャート。

【図4】同実施例の動作を説明するためのフローチャート。

【図5】同実施例に係わる予備ディスクドライブの記憶内容を説明するための概念図。

【図6】同実施例に係わる予備ディスクドライブの記憶内容を説明するための概念図。

(8)

特開平7-56693

13

14

【図7】同実施例の動作を説明するためのフローチャート。

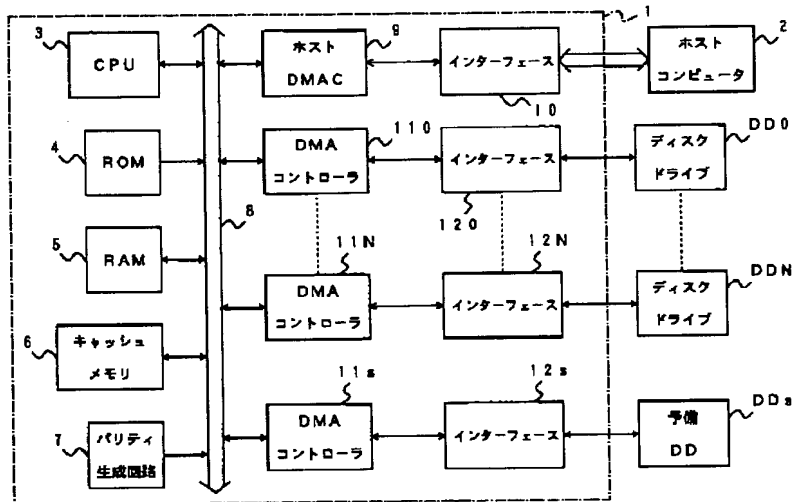
【図8】従来のレベル5のディスク・アレイ装置の構成を説明するための概念図。

【符号の説明】

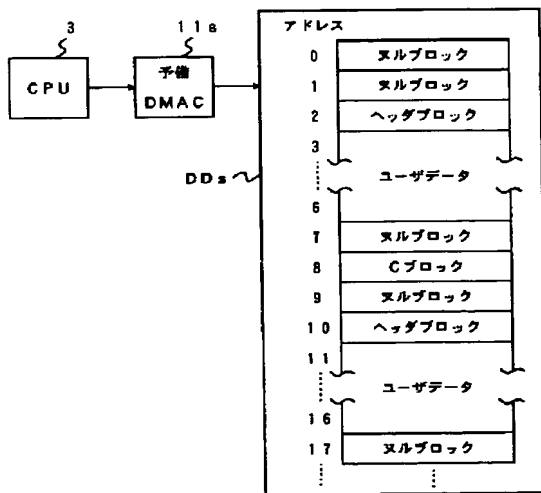
1…ディスク・アレイ・コントローラ、2…ホストコン

ピュータ、3…CPU、6…キャッシュメモリ、7…パリティ生成回路、9…ホストDMAコントローラ、DD0～DDN…ディスクドライブ、110～11N…ディスクDMAコントローラ、DDs…予備ディスクドライブ、11s…予備ディスクDMAコントローラ。

【図1】



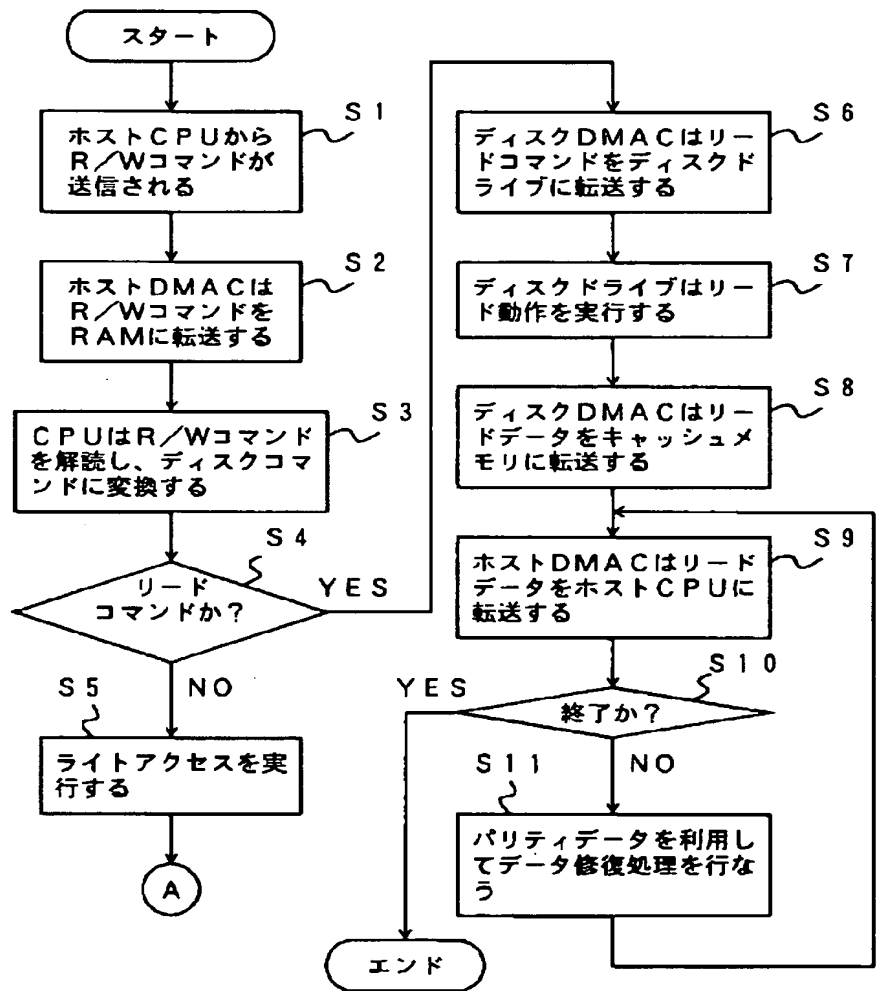
【図5】



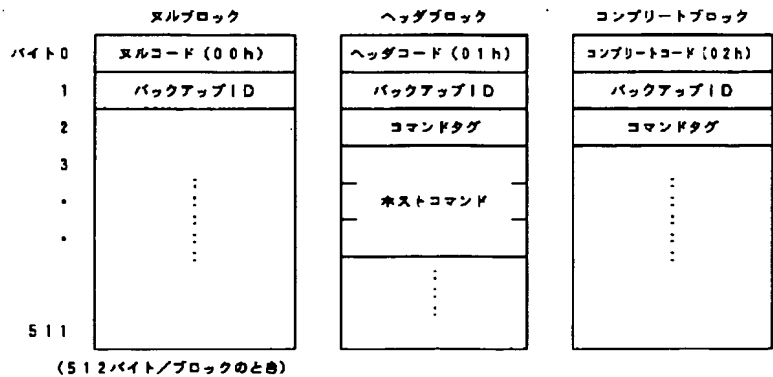
(9)

特開平7-56693

【図2】



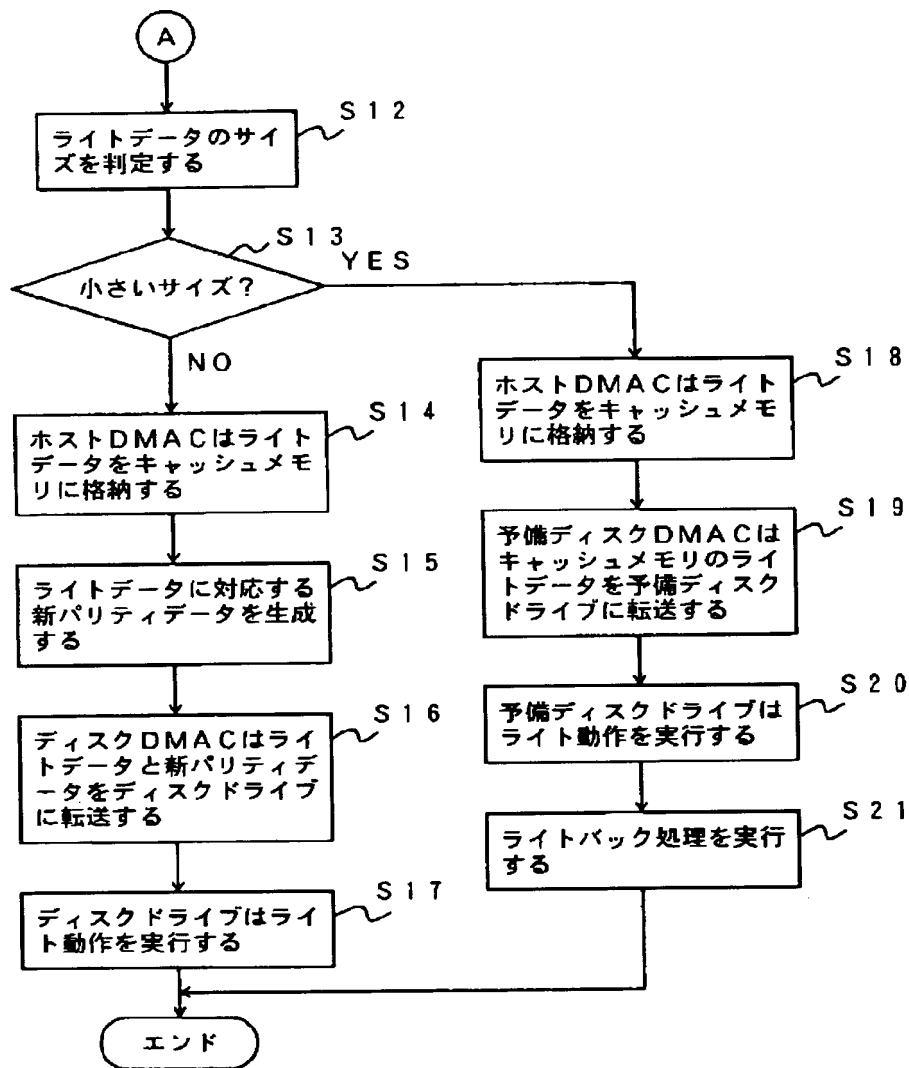
【図6】



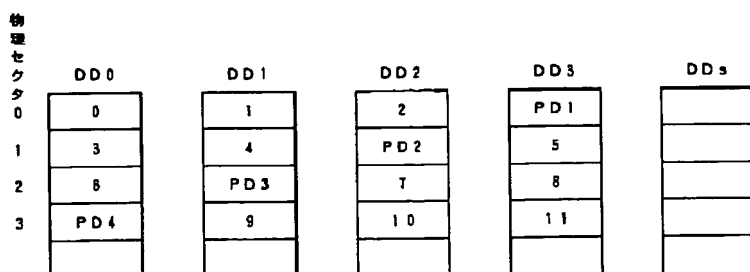
(10)

特開平7-56693

【図3】



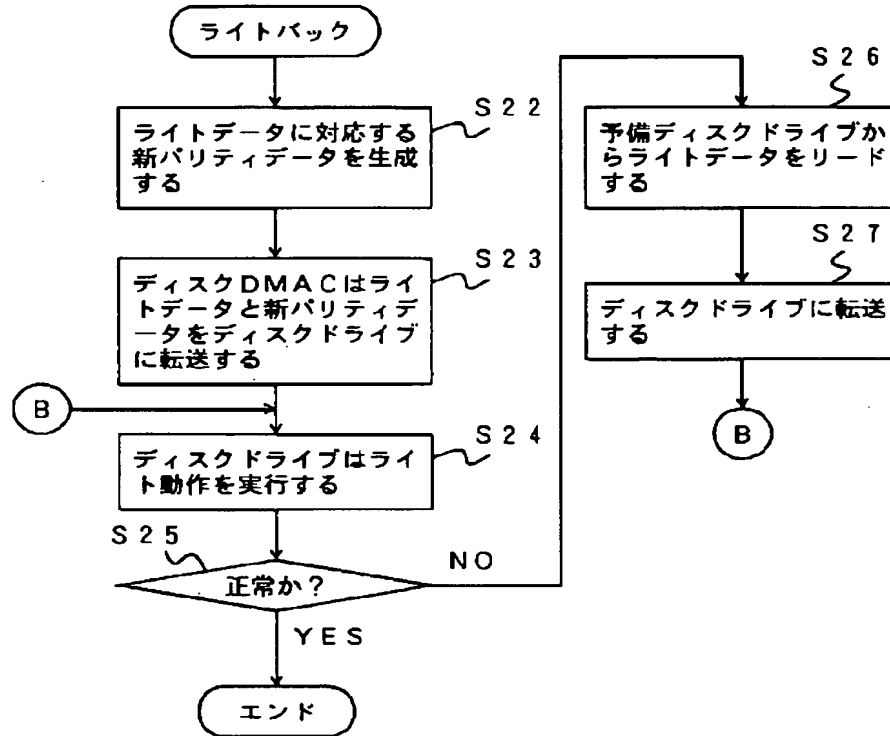
【図8】



(11)

特開平 7-56693

【図4】



(12)

特開平 7-56693

【図7】

